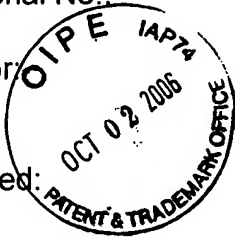


IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Schroeder et al.

Serial No.: 10/549,246

For:



Determining the Quality of Biomolecule Samples

Filed: June 7, 2006

Examiner: Not Yet Assigned

Art Unit: 2624

Customer No.: 27623

Confirmation No.: 2446

Attorney Docket: 20030461-3

COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, VA 22313-1450

**REQUEST FOR ENTRY OF PRIORITY CLAIM
AND SUBMISSION OF PRIORITY DOCUMENT**

Dear Sir:

Applicant hereby requests that a priority claim under 35 U.S.C. §119 be entered in the above-identified application as follows: **German** Application No. **103 15 581.3** filed on **April 5, 2003**, for the above noted application.

We are also enclosing a certified copy of the priority document; **German** Application No. **103 15 581.3** filed **April 5, 2003**, for filing in the above noted application.

It is respectfully requested that this application be passed to allowance.

Respectfully submitted,

Date: September 28, 2006

Paul D. Greeley
Attorney for Applicants
Registration No. 31,019
Ohlandt, Greeley, Ruggiero & Perle, L.L.P.
One Landmark Square, 10th Floor
Stamford, Connecticut 06901-2682
Telephone: (203) 327-4500
Telefax: (203) 327-6401

BEST AVAILABLE COPY

BUNDESREPUBLIK DEUTSCHLAND



Prioritätsbescheinigung DE 103 15 581.3 über die Einreichung einer Patentanmeldung

Aktenzeichen: 103 15 581.3

Anmeldetag: 05. April 2003.

Anmelder/Inhaber: Agilent Technologies Inc. (n.d.Ges.d. Staates Delaware), Palo Alto, Calif./US

Bezeichnung: Verfahren zur Qualitätsbestimmung von RNA-Proben

IPC: C 12 Q 1/68

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

München, den 25. Juli 2006
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag


Wallner

BEST AVAILABLE COPY

Verfahren zur Qualitätsbestimmung von RNA-Proben

Die Erfindung betrifft ein Verfahren zur Qualitätsbestimmung von RNA-Proben nach dem Oberbegriff des Anspruchs 1.

Ribonukleinsäuren, in der englischen Sprache ribonucleic acids, abgekürzt RNA, sind in den Zellen aller Lebewesen und auch in einigen Viren enthaltene Biopolymere, die in verschiedener Form vor allem an der Übersetzung der genetischen Information von DNA in Proteine beteiligt sind. Daher wird RNA für die Untersuchungen des Genoms in MicroArray-Hybridisierungs-Experimenten und RT-PCR, auch beispielsweise für Northern Blots, RNase-Protection-Assays oder cDNA-Synthese herangezogen. Das Resultat dieser Experimente und die Signifikanz der Ergebnisse sind in hohem Maße von der Qualität der verwendeten RNA-Proben abhängig.

Die Größenverteilung von RNA-Biopolymeren dient als Maßstab für die Intaktheit und damit die Qualität der vorliegenden RNA-Probe. Die RNA-Größenverteilung variiert in Abhängigkeit von der Herkunft des Materials und der Präparationsmethode, wird jedoch maßgeblich durch Kontamination mit RNA-degradierenden Enzymen (RNasen) oder durch mechanische Scherkräfte bei unsachgemäßer Handhabung beeinflusst. In jedem Fall beobachtet man bei einer Degradation eine Verschiebung der RNA-Polymerlängen zu kleineren Größen.

RNA Biopolymere zeichnen sich dadurch aus, dass ihre Ladung nahezu proportional zu ihrer Länge ist. Eine Elektrophorese, d.h. eine Trennung von Molekülen nach ihrem Ladung/Masse-Verhältnis in einem Substrat durch Anlegen eines elektrischen Feldes, eignet sich daher für die Analyse der Größenverteilung einer RNA-Probe. Lab-on-a-Chip-Analysen mit dem Agilent 2100 Bioanalyzer bieten ein hochreproduzierbares und hochauflösendes elektrophoretisches Verfahren. So gewonnene digital vorliegende Elektropherogramme bilden daher eine ideale Ausgangsbasis für die weitere Verarbeitung.

BEST AVAILABLE COPY

Eine sichere Qualitätsbestimmung ist bislang nur manuell und daher mehr oder weniger subjektiv möglich. Jede einzelne Probe muss von einem erfahrenen Biochemiker auf Intaktheit überprüft werden, bevor sie in weiteren Experimenten verwendet werden darf. Mit wachsender Anzahl der RNA-Experimente und neuer Hochdurchsatztechnik wird eine manuelle Qualitätsbestimmung untragbar.

Es sind Verfahren bekannt, bei denen die Qualität aufgrund eines oder nur weniger Merkmale abgeschätzt wird. Der bislang beste Ansatz legt ein Kriterium für das Flächenverhältnis des 28S- und des 18S-rRNA-Fragments fest. Intakte RNA-Proben weisen theoretisch einen Wert von 2 auf. Mit fortschreitender Degradierung wird das Verhältnis immer kleiner. Es bietet eine erste, jedoch relativ ungenaue Möglichkeit zwischen intakter und degradierter RNA zu unterscheiden, da in der Praxis oft starke Abweichungen des Verhältnisses vom theoretischen Wert auftreten. Dieses Kriterium wird als Stand der Technik für eine automatische Qualitätsbestimmung der RNA-Proben angesehen.

Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren zur automatischen und zuverlässigen Qualitätsbestimmung anhand von Elektropherogrammen zu entwickeln. Das Verfahren soll unabhängig von der Quelle und Art des RNA-Materials effektiv einsetzbar sein, das heißt unabhängig von der biologischen Spezies, dem Zellzustand, dem Gewebe- bzw. Organtyp, dem Organismus, der Konzentration und der Präparationsmethode der RNA-Proben.

Die Erfindung und ihre Vorteile

Das erfindungsgemäße Verfahren mit den Merkmalen des Anspruchs 1 hat gegenüber dem Stand der Technik den Vorteil, dass die RNA-Proben mit einem objektiven, einheitlichen und reproduzierbaren Qualitätswert charakterisiert werden können. Dies eröffnet neue Möglichkeiten für die Qualitätskontrolle und Qualitätssicherung, wie beispielsweise objektive Qualitätsvergleiche von RNA-

BEST AVAILABLE COPY

eine Präparation der zellulären Gesamt-RNA, bestehend aus mRNA, rRNA sowie RNA.

Es besteht außerdem die Möglichkeit, das Verfahren mit anderen RNA-Assays des Agilent 2100 Bioanalyzer Systems durchzuführen. Bei dem „Eukaryote Total RNA Pico-Assay“ werden beispielsweise RNA-Konzentrationen im Picogramm-Bereich eingesetzt. Das Verfahren kann ebenfalls für die Qualitätsbestimmung der RNA aus Prokaryonten eingesetzt werden. Prokaryonten-RNA unterscheidet sich von der Eukaryonten-RNA hauptsächlich in den auftretenden Polymerlängen der ribosomalen Fragmente. Schließlich kann das Verfahren für mRNA-Assays (Eukaryote mRNA Nano, Eukaryote mRNA Pico, Prokaryote mRNA Nano, Prokaryote mRNA Pico) verwendet werden. mRNA-Präparationen enthalten idealerweise ausschließlich den mRNA-Anteil der zellulären Gesamt-RNA.

Als Elektropherogramme werden Diagramme einer Elektrophorese bezeichnet. In den Diagrammen wird die Quantität der gemessenen RNA-Fragmente gegen ihre Migrationsszeit aufgetragen. Diese können beispielsweise mit Agilent 2100 Bioanalyzer oder mit klassischen Verfahren der Gel-Elektrophorese bestimmt werden. Umgangssprachlich wird auch die Gesamtheit der dem Elektropherogramm zugrundeliegenden Datenpunkte als Elektropherogramm bezeichnet.

Die Datenpunkte eines Elektropherogramms bilden die Eingabe für das Verfahren. Das Verfahren extrahiert im ersten Schritt wenige vorgegebene Merkmale (f_1, \dots, f_n) aus den Elektropherogrammen. Im zweiten Schritt wird aus diesen Merkmalen der Qualitätswert mit Hilfe eines Qualitätsalgorithmus berechnet.

BEST AVAILABLE COPY

Nach einer vorteilhaften Ausgestaltung der Erfindung wird der Qualitätsalgorithmus durch folgende Verfahrensschritte bestimmt:

- A. Anlegen einer statistisch signifikanten Versuchsmenge von RNA-Elektropherogrammen zu einer vorgegebenen Menge von RNA-Proben,
- B. Vorgabe eines Qualitätskennzeichens q zu jedem Elektropherogramm,
- C. Extraktion möglichst vieler aussagekräftiger Merkmale aus den Elektropherogrammen mittels Methoden der Datenanalyse,
- D. Bestimmung funktionaler Zusammenhänge zwischen Qualitätskennzeichen und bestimmten Merkmalskombinationen, beispielsweise mit einem adaptiven Verfahren,
- E. Zuordnung eines Gütwertes zu jedem der funktionalen Zusammenhänge, beispielsweise die nach der Bayes'schen Methode ermittelte a posteriori Wahrscheinlichkeit,
- F. Bestimmung des funktionalen Zusammenhangs mit dem höchsten Gütwert als Qualitätsalgorithmus.

Als Versuchsmenge werden möglichst viele Elektropherogramme angelegt. Es ist besonders wichtig, dass die Versuchsmenge die Daten realer Anwendungen widerspiegelt.

Alle Proben werden im Vorfeld sorgfältig mit einem Qualitätskennzeichen versehen. Dieses Qualitätskennzeichen ist die Zielgröße, die für die spätere Auswahl der besten Merkmalskombination und das Trainieren des neuronalen Netzes verwendet wird.

Die Qualität einer RNA-Probe ist eine kontinuierliche Größe, so dass sich keine natürlichen Qualitätsklassen ergeben. Daher werden nach einer weiteren vorteilhaften Ausgestaltung der Erfindung diskrete Qualitätsklassen festgelegt. Beispielsweise können sieben Qualitätsklassen eingeführt werden. Die qualitativ schlechtesten RNA-Proben werden mit dem Qualitätskennzeichen „1“ versehen und in die erste Klasse eingeordnet. Die qualitativ etwas besseren RNA-Proben werden mit „2“ gekennzeichnet und in die zweite Klasse eingeordnet usw. Die qualitativ besten RNA-Proben enthalten schließlich das Kennzeichen „7“.

BEST AVAILABLE COPY

Der adaptive Ansatz hat den Vorteil, dass die beste Merkmalskombination für die Qualitätsbestimmung automatisch ausgewählt wird und auf dieser Merkmalskombination die Qualität adaptiv gelernt wird.

In diesem Punkt stellt die Gesamtheit aller digitalen Elektropherogramme mit zugehörigen Qualitätskennzeichen $q \in \{1, \dots, 7\}$ und die komplette Wissensbasis für die Weiterentwicklung des Verfahrens dar.

Das Ziel der Extraktion von Merkmalen aus den Elektropherogrammen ist es, aus dem Elektropherogramm möglichst viele aussagekräftige Merkmale zu extrahieren. Findungsgemäß teilt man hierzu das Elektropherogramm in folgende Bereiche auf: Pre-Region, Marker-Region, 5S-Region, Fast-Region, 18S-Region, Inter-Region, 28S-Region und Post-Region.

Jeder dieser Bereiche wird anschließend getrennt betrachtet und liefert einige bereichsspezifische, sogenannte lokale Merkmale, die gemeinsam die Form des Elektropherogramms in dem betroffenen Bereich ausreichend genau beschreiben. Außerdem werden einige globale, d.h. bereichsübergreifende, Merkmale extrahiert. Das Ergebnis dieses Verfahrensschrittes ist eine Liste von beispielsweise ca. 100 Merkmalen pro Elektropherogramm.

Die Basis der Datenanalyse bildet eine Liste mit den im Elektropherogramm erkannten Maxima, die als Peaks bezeichnet werden. Die Erkennung der Peaks wird durch Integration der Datenkurve erreicht. Die Integration liefert neben der Position der Peaks auch die Start- und Endpunkte der Peaks sowie deren Höhe, Breite und Fläche.

In Anlehnung an die Agilent 2100 Bioanalyzer System Software werden einige Peaks als „Ladder Peak“, „Marker“, „18S-Peak“- und „28S-Peak“ markiert. Das neue Verfahren zur Integration und Markierung weist gegenüber dem Stand der Technik eine deutliche Weiterentwicklung und Verbesserungen der Genauigkeit und der Robustheit gegen Anomalien wie „Ghost Peak“ und „Spikes“ auf. Gemäß der Erfindung stellt man dafür ein statistisches lineares Modell über Position, Höhe und

BEST AVAILABLE COPY

ne der ersten vier Ladder Peaks bereit. Die vier Peaks in der Ladder-, die einsam am besten zum Modell passen, werden als Ladder Peaks bezeichnet markiert.

Der erste Ladder Peak in der Ladder ist der Lower-Marker, dessen Position, Höhe und Fläche bis auf den Drift-Effekt mit den Positionen, Höhen bzw. Flächen der anderen Marker der restlichen Proben eines Chips übereinstimmen. Wieder wird ein statistisches Modell aufgestellt, das diesmal neben der Position, Höhe und Fläche auch den Lower-Marker-Verlauf in den Proben eines Chips berücksichtigt. Die 13 Peaks, ein Peak aus jeder Probe eines Chips, die am besten zu diesem Modell passen, werden als Lower-Marker markiert.

Schließlich werden der Zusammenhang zwischen den Positionen des Markers und der 18S- und 28S-Peaks in einem Modell zusammengefasst und die entsprechenden Peaks als 18S- und 28S-Peaks markiert. Bei stark degradierten RNA-Proben sind 18S und 28S-Peaks nicht mehr von dem Hintergrund unterscheidbar. In diesen Fällen wird die geschätzte Position der 18S- und 28S-Peaks berechnet, um die nachfolgende Aufteilung in Bereiche für alle Qualitätsklassen zu ermöglichen.

Die auf diese Weise erhaltene Markierung weicht lediglich in 0.8% der Fälle von der manuellen Markierung der „Lower Marker“ und in 1.2% der Fälle von der manuellen Markierung der 18S- und 28S-Peaks ab.

In weiterer Ausbildung der Erfindung erfolgt auf Basis der Markierung eine Aufteilung jedes Elektropherogramms in die oben erwähnten acht aneinander angrenzenden Bereiche, die den gesamten Datenbereich abdecken. Der Bereich vor dem Lower-Marker wird als Pre-Region bezeichnet. Die Marker-Region deckt sich mit dem Bereich, den der Lower-Marker-Peak einnimmt. Die 18S- und 28S-Regionen erstrecken sich jeweils über den 18S-Peak bzw. 28S-Peak. Zwischen Marker-Region und 18S-Region liegen zwei Regionen, die 5S-Region und die Fast-Region. Die ungefähre Grenze zwischen diesen beiden Bereichen wird aus der Position des Lower Markers und der 5.8S / 5S / tRNA Peaks in den Proben mit vorhandener 5.8S und 5S rRNA sowie tRNA bestimmt und anhand der jeweiligen Position des Lower

BEST AVAILABLE COPY

Marker-auffälligen Proben übertragen. Die Inter-Region liegt zwischen den 18S- und 28S-Bereichen. Figur 1 der Zeichnung veranschaulicht die vorgenommene Aufteilung.

Die Korrektur der Basislinie im Elektropherogramm, auch genannt Baseline, wird ebenfalls in Anlehnung an Agilent 2100 Bioanalyzer System Software mit wesentlichen Verbesserungen vorgenommen. In den Bereichen Pre-Region und Post-Region verläuft die Baseline bis auf das Rauschen idealerweise auf einem konstanten Niveau. Das Niveau kann von Elektropherogramm zu Elektropherogramm sehr unterschiedliche Werte annehmen. In einigen Fällen kann die Baseline auch eine Steigung oder gar Wellen aufweisen. Letzteres stellt ein deutliches Indiz für ein aufgetretenes Problem während der Datenakquisition dar.

Der Grundgedanke der Baseline-Korrektur besteht darin, den konstanten, oder den mit der Zeit proportionalen zu- bzw. abnehmenden Hintergrundanteil aus dem Datensignal zu entfernen. Dazu wird erfindungsgemäß versucht, eine Gerade zu finden, die in den Bereichen Pre-Region und Post-Region mit dem Datensignal bis auf das Rauschen übereinstimmt, d.h. die im Mittel um Noise Standard Deviation σ_{noise} von dem Datensignal abweicht. Für die Berechnung der Noise Standard Deviation σ_{noise} wird die übliche aus der Literatur bekannte Formel benutzt.

Vor der eigentlichen Merkmalsextraktion wird das Datensignal auf das globale Maximum in der 5S-Region, der Fast-Region, der 18S-Region, der Inter-Region und der 28S-Region normiert. Die Marker-Region wird hier außer Acht gelassen, um auf diese Weise verschiedene Konzentrationen besser zu handhaben.

Neben der ursprünglichen Datenkurve verwendet man weitere geglättete Datenkurven. Zur Glättung der Datenkurve verwendet man vorzugsweise den Savitzky-Golay-Filter und den Rollingball-Algorithmus nach EP 0 969 283 A1.

Für alle Regionen bieten sich folgende lokale Merkmale zur Extraktion aus der ursprünglichen und den geglätteten Datenkurven an:

- Minimaler/ Maximaler Wert im Bereich